

BAB 2

TINJAUAN PUSTAKA DAN DASAR TEORI

2.1 Tinjauan Pustaka

2.1.1 Penelitian Terkait

Penelitian ini menggunakan beberapa pustaka yang berkaitan dengan penelitian ini. Hal ini berfungsi untuk pedoman dan pembanding penelitian yang akan dilakukan.

Adapun perbandingan tinjauan pustaka tersebut dapat dilihat pada tabel perbandingan penelitian Sistem Prediksi Kelulusan Mahasiswa Berbasis Web Menggunakan Metode Algoritma *Naive Bayes*.

Tabel 2.1 Penelitian Terkait

Nama	Objek Penelitian	Metode	Atribut
Amalia, Naziah, dkk. (2010)	Universitas Telkom	<i>Naïve Bayes Classifier</i>	Kota asal, sks, IPK, Kelompok Keahlian, TA1, dan TA2.
Hananto, Valentinus Roby (2017)	Prodi Sistem Informasi – S1 Stikom Surabaya	<i>Naïve Bayes Multilayer Perceptron, SMO, J48 dan REPTree</i>	Dosen Wali, IPK, SKSK (SKS Kumulatif), STS (Status Mahasiswa), STS_Tempuh_KP (Status menempuh kerja praktik), dosen1 (dosen pembimbing 1),

			dosen 2 (dosen pembimbing 2) dan Status tepat waktu (ya/tidak).
Nugroho, Mohamad Fajarianditya dan Wibowo, Setyoningsih (2017)	Fakultas Ilmu Komputer UNAKI Semarang	<i>Naïve Bayes Classifier</i>	Status pekerjaan dan IPK semester 4.
Mauriza, Ahmad Fikri (2014)	Fakultas Komunikasi Dan Informatika UMS	<i>Naïve Bayes</i>	Jurusan asal sekolah, gender, asal sekolah, rata SKS, rata MK, asisten Lab
Yulianti (2018)	Program Studi Sarjana STMIK Akakom Yogyakarta	<i>K- Nearest Neighbor</i>	Indeks prestasi semester 1, indeks prestasi kumulatif sampai semester 4, total sks sampai semester 4, dan jurusan asal sekolah, jumlah nilai D dalam sks, jumlah nilai E dalam sks.
Nugroho Tri Atmojo (2021)	Program Studi Informatika STMIK AKAKOM Yogyakarta	<i>Naïve Bayes</i>	Indeks prestasi semester 1 sampai dengan semester 4, IPK, jumlah semester tidak aktif dan status lulus.

Amalia, Naziah, dkk. (2017) melakukan prediksi ketepatan waktu lulus mahasiswa menggunakan algoritma *Naïve Bayes Classifier* dan atribut yang digunakan adalah kota asal, sks, IPK, Kelompok Keahlian, TA1, dan TA2. Kelas targetnya yaitu tepat waktu dengan 2 klasifikasi ya dan tidak dan diperoleh tingkat akurasi sebesar 86%.

Hananto, Valentinus Roby (2017) melakukan perbandingan akurasi antar metode data mining untuk memprediksi kelulusan mahasiswa sebagai penunjang angka efisiensi edukasi. Metode yang digunakan yaitu *Naïve Bayes*, *Multilayer Perceptron*, *SMO*, *J48*, dan *REPTree*. Hasil yang diperoleh dari pengujian bahwa metode *Naïve Bayes* mempunyai tingkat akurasi tertinggi dan *error rate* terkecil dengan tingkat akurasi 57,3% menggunakan tahapan CRISP-DM. Atribut yang digunakan adalah Dosen Wali, IPK, SKS Kumulatif, Status Mahasiswa(Tugas Akhir), Status menepuh Kerja Praktik, Dosen Pembimbing1,dan Dosen Pembimbing2. Kelas target yang digunakan adalah Lulus Tepat Waktu yang terdiri dari 2 klasifikasi yaitu Ya dan Tidak.

Nugroho, Mohamad Fajarianditya dan Wibowo, Setyoningsih (2017) melakukan seleksi *forward selection* untuk menentukan atribut yang berpengaruh pada klasifikasi kelulusan mahasiswa menggunakan algoritma *Naïve Bayes* dan diperoleh kesimpulan bahwa atribut yang berpengaruh dalam penentuan kelulusan mahasiswa tepat waktu adalah status pekerjaan dan IPK semester 4 dengan akurasi 97,14% .

Mauriza, Ahmad Fikri (2014) melakukan prediksi kelulusan mahasiswa menggunakan metode *Naïve Bayes* dan atribut yang digunakan yaitu jurusan asal sekolah, gender, daerah asal sekolah, rata SKS, rata MK, asisten Lab dan untuk klasifikasinya adalah tepat apabila memiliki nilai lulus ≤ 4 tahun dan terlambat apabila memiliki nilai lulus > 4 tahun.

Yulianti, S. (2018) melakukan implementasi data mining untuk memperkirakan masa studi mahasiswa dengan metode *K- Nearest Neighbor (KNN)* dan atribut yang digunakan yaitu indeks prestasi semester 1 sampai 4 dan jumlah sks dari semester 1 sampai semester 4 dengan tingkat akurasi sebesar 70 %.

Selanjutnya skripsi yang dibuat adalah sistem prediksi kelulusan mahasiswa dengan menggunakan metode *Naive Bayes* dengan attribut yang digunakan adalah IP Semester 1 sampai dengan 4, IPK, jumlah semester tidak aktif dan status lulus.

2.2 Dasar Teori

2.2.1 Sekilas tentang Program Studi Informatika STMIK AKAKOM

Program Studi Informatika merupakan salah satu cabang dari pohon ilmu Informatika dan Komputer, yang turut mengembangkan bidang ilmu teknologi informasi. Adapun penekanannya pada aspek penyimpanan, pengolahan, dan penyampaian informasi secara algoritmik dengan bantuan komputer. Tiga (3) paradigma yang digunakan adalah: paradigma teori, paradigma abstraksi, dan paradigma perancangan. Ketiga paradigma ini saling melengkapi dan kait-mengait antara satu dengan yang lain. Secara umum ketiga paradigma tersebut (teori,

abstraksi, perancangan) merupakan prinsip dasar dari masing-masing bidang ilmu yang diwakili matematika, ilmu alam, teknik.

Jumlah alumni Program Studi Informatika hingga saat ini mencapai sekitar 1000 mahasiswa, yang telah bekerja pada perbankan, perminyakan, pendidikan, kesehatan, dan bidang-bidang lain. Banyak mahasiswa yang telah bekerja sebelum lulus dari pendidikannya. Hal ini membuktikan bahwa ilmu yang diterima mahasiswa bisa digunakan kapan saja. Beberapa lapangan kerja alumni diantaranya Krakatau Steel, BRI, Pertamina, TOTAL, (SGM), Instansi Negeri (Kejaksanaan, Pemda, PTN), dan Perguruan Tinggi Swasta.

2.2.2 Algoritma *Naive Bayes*

Algoritma *Naive Bayes* merupakan salah satu algoritma yang terdapat pada teknik klasifikasi. *Naive Bayes* merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris *Thomas Bayes*, yaitu memprediksi peluang di masa depan berdasarkan pengalaman dimasa sebelumnya sehingga dikenal sebagai *Teorema Bayes*. Teorema tersebut dikombinasikan dengan *Naive* dimana diasumsikan kondisi antar atribut saling bebas. Klasifikasi *Naive Bayes* diasumsikan bahwa ada atau tidak ciri tertentu dari sebuah kelas tidak ada hubungannya dengan ciri dari kelas lainnya.

Persamaan dari teorema *Bayes* adalah :

$$P(C_i | X) = \frac{p(X | C_i)P(C_i)}{p(X)} \quad \dots \text{Persamaan (1)}$$

Keterangan :

X : Data dengan *class* yang belum diketahui

C_i : Suatu variabel yang harus dideskripsikan secara probabilistik

$P(C_i|X)$: Probabilitas hipotesis C_i berdasarkan kondisi X (*posteriori probability*)

$P(C_i)$: Probabilitas hipotesis C_i (*prior probability*)

$P(X|C_i)$: Probabilitas X berdasar kondisi pada hipotesis C_i

$P(X)$: Probabilitas dari X

Berikut ini merupakan contoh *data training* sebanyak 14 data dengan *output* main sepak bola atau tidak. Setiap data ditandai dengan atribut cuaca, temperatur, kelembaban, dan angin yang dapat dilihat pada Tabel 2.2

Tabel 2.2 *Data Training* Cuaca dan Keputusan Main atau Tidak

Cuaca x1	Temperatur x2	Kelembaban x3	Angin x4	Main atau Tidak y
Cerah	Panas	Tinggi	Kecil	Tidak
Cerah	Panas	Tinggi	Besar	Tidak
Mendung	Panas	Tinggi	Kecil	Ya
Hujan	Sedang	Tinggi	Kecil	Ya
Hujan	Dingin	Normal	Kecil	Ya
Hujan	Dingin	Normal	Besar	Tidak
Mendung	Dingin	Normal	Besar	Ya
Cerah	Sedang	Tinggi	Kecil	Tidak
Cerah	Dingin	Normal	Kecil	Ya
Hujan	Sedang	Normal	Kecil	Ya
Cerah	Sedang	Normal	Besar	Ya
Mendung	Sedang	Tinggi	Besar	Ya
Mendung	Panas	Normal	Kecil	Ya

Hujan	Panas	Tinggi	Besar	Tidak
Cerah	Dingin	Tinggi	Besar	?

Berikut ini merupakan contoh perhitungan metode *Naive Bayes* yang menggunakan rumus pada persamaan (1) untuk menentukan kelas dari data baru berikut:

(Cuaca = cerah, Temperatur = dingin, Kelembaban = tinggi, Angin = besar)

Langkah – langkah perhitungan yang dikerjakan sebagai berikut:

a. Mencari *probabilitas prior*

Probabilitas prior ini menyatakan berapa peluang munculnya keputusan main sepak bola $P(C1)$ dan peluang keputusan tidak main sepak bola $P(C2)$. Misalkan N adalah jumlah total keputusan main sepak bola dan tidak, $N1$ menyatakan jumlah keputusan main sepak bola dan $N2$ menyatakan jumlah keputusan tidak main sepak bola. Berdasarkan pada Tabel 2.2 jumlah total keputusan main sepak bola dan tidak sebanyak 14, jumlah keputusan main sepak bola sebanyak 9, dan jumlah keputusan tidak main sepak bola sebanyak 5. Didapatkan rumus sebagai berikut :

$$\begin{aligned}
 P(C1) &= N1/N & P(C2) &= N2/N \\
 &= 9/14 & &= 5/14 \\
 &= 0,64 & &= 0,36
 \end{aligned}$$

b. Mencari *Probabilitas bersyarat (likelihood)*

Probabilitas bersyarat (likelihood), $P(X/C_i)$ ini menyatakan peluang munculnya X jika diketahui C_i , misalkan:

X : (Cuaca=cerah, Temperatur = dingin, Kelembaban = tinggi, Angin= besar)

C_i : C_1 adalah main sepak bola dan C_2 adalah tidak main sepak bola.

Sehingga *likelihood* dapat dihitung dengan mengalikan hasil dari masing – masing nilai probabilitas per-atribut yang berdasarkan pada C_i .

Langkah – langkah menghitung *likelihood* :

- Menghitung probabilitas per-atribut

Misalkan menghitung probabilitas Angin = besar dengan C_i adalah C_1 yaitu main sepak bola. Berdasarkan Table 2.2 banyaknya Angin = besar yang keputusannya adalah main sepak bola ada 3 dan banyaknya kelas dengan keputusan main sepak bola ada 9. Maka probabilitasnya adalah

$$\begin{aligned}
 P(\text{Angin=Besar}|\text{main}) &= \frac{\text{jumlah Angin Besar keputusan main}}{\text{jumlah seluruh keputusan main}} \\
 &= 3/9 \\
 &= 0,33
 \end{aligned}$$

Melakukan perhitungan dengan cara yang sama pada probabilitas per-atribut yang lain, sehingga diperoleh hasil :

$$P(\text{Cuaca}=\text{Cerah}/\text{Main}) = 2/9 = 0,22$$

$$P(\text{Cuaca}=\text{Cerah}/\text{Tidak}) = 3/5 = 0,60$$

$$P(\text{Temperatur}=\text{dingin} / \text{Main}) = 3/9 = 0,33$$

$$P(\text{Temperatur}=\text{dingin} / \text{Tidak}) = 1/5 = 0,20$$

$$P(\text{kelembaban}=\text{tinggi} / \text{Main}) = 3/9 = 0,33$$

$$P(\text{kelembaban}=\text{tinggi} / \text{tidak}) = 4/5 = 0,80$$

$$P(\text{Angin}=\text{Besar} / \text{Main}) = 3/9 = 0,33$$

$$P(\text{Angin}=\text{Besar} / \text{Tidak}) = 3/5 = 0,60$$

- Mengalikan semua probabilitas per-atribut yang berdasarkan pada C_i yaitu C_1 adalah main yang disebut *Likelihood Ya* atau C_2 adalah tidak main yang disebut *Likelihood Tidak*.

- *Likelihood Ya*

$$P(X/\text{Main}) = P(\text{Cuaca}=\text{Cerah}/\text{Main}) * P(\text{Temperatur}=\text{dingin}/\text{Main}) * P(\text{kelembaban}=\text{tinggi}/\text{Main}) * P(\text{Angin}=\text{Besar} / \text{Main})$$

$$= 0,22 * 0,33 * 0,33 * 0,33$$

$$= 0,0080$$

- *Likelihood Tidak*

$$P(X/Tidak) = P(Cuaca=Cerah/Tidak) * P(Temperatur=dingin/Tidak) * P(kelembaban=tinggi /Tidak) * P(Angin=Besar /Tidak)$$

$$= 0.60 * 0,20 * 0,80 * 0,60$$

$$= 0.0576$$

- c. Mengalikan *Likelihood* dengan *Prior*

Perkalian *Likelihood* dengan *Prior* merupakan hal yang penting untuk menemukan *posterior*.

- Perkalian *Likelihood* dengan *Prior* pada keputusan Main

$$= P(X/C_i) * P(C_i)$$

$$= P(X/Main) * P(Main)$$

$$= 0,0080 * 0,64$$

$$= 0,00512$$

- Perkalian *Likelihood* dengan *Prior* Tidak Main

$$= P(X/C_i) * P(C_i)$$

$$= P(X/Tidak) * P(Tidak)$$

$$= 0,0576 * 0,36$$

$$= 0,020736$$

d. Menghitung *probabilitas posterior*

Probabilitas posterior, $P(C_i|X)$ ini menyatakan probabilitas keluarnya hasil C_i jika diketahui nilai X tertentu. *Probabilitas posterior*, $P(C_i|X)$ dicari dengan menggunakan rumus pada persamaan (1) :

- Probabilitas *Posterior Main*

$$\begin{aligned}
 P(C_i | X) &= \frac{p(X | C_i)P(C_i)}{p(X)} \\
 \text{posterior} &= \frac{\text{likelihood} * \text{prior}}{\text{evidence}} \\
 &= \frac{p(X | \text{Main}) * P(\text{Main})}{P(X | \text{Main}) * P(\text{Main}) + P(X | \text{Tidak}) * P(\text{Tidak})} \\
 &= \frac{0,00512}{0,00512 + 0,020736} \\
 &= 0,198019802
 \end{aligned}$$

- Probabilitas *Posterior Tidak Main*

$$\begin{aligned}
 P(C_i | X) &= \frac{p(X | C_i)P(C_i)}{p(X)} \\
 \text{posterior} &= \frac{\text{likelihood} * \text{prior}}{\text{evidence}} \\
 &= \frac{p(X | \text{Tidak}) * P(\text{Tidak})}{P(X | \text{Tidak}) * P(\text{Tidak}) + P(X | \text{Main}) * P(\text{Main})} \\
 &= \frac{0,020736}{0,020736 + 0,00512} \\
 &= 0,801980198
 \end{aligned}$$

Nilai *probabilitas posterior* Tidak Main sebesar 0,801980198 lebih besar dari nilai *probabilitas posterior* Main yang bernilai 0,198019802 dan nilai *probabilitas possterior* Tidak Main mendekati nilai 1, sehingga dengan *Naive Bayes* prediksi dapat disimpulkan Tidak Main untuk data *input* ini yang berdasarkan pada estimasi *probabilitas* yang dipelajari dari *data training*.